REVERTIBLE DEEP CONVOLUTIONAL NETWORKS WITH ITERATED DIRECTIONAL FILTER BANK

Can Xu *Hongkai Xiong*[*]

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

Efficient image representation methods are of great significance in image processing. This paper proposes an invertible deep convolutional network where the entire architecture is constructed by a decomposition process and a frequency recombination process. The decomposition allows frequency division in both the lower and higher part at the same time on each layer. It is implemented by using iterated directional filter banks. Perfect reconstruction is available if we adopt biorthogonal qincunx filter banks. The remaining frequency recombination process is adjoined to the architecture to transform the uniform frequency partition to nonuniform one and thereby increase the efficiency and flexibility. Numerical experiments reveal that the underlying network has good performance especially for images with large amount of fine-grained information.

Index Terms— Deep convolutional network, revertible, directional filter banks, nonlinear approximation, denoising

1. INTRODUCTION

With the occurrence of wavelet, the developments of signal processing have taken a big step forward [1]. The ability to efficiently approximate signals containing point-wise singularities makes wavelet analysis a powerful tool of representing one-dimensional piecewise smooth functions than Fourier analysis. However, it becomes not so powerful when applied to two or higher dimensional signals.

To this point, multiscale geometric analysis (MGA) which aims to construct a new optimal representation of high-dimensional function become a significant subject for researchers. With the development of MGA [2-4], several image representation methods have been developed to capture the geometric regularity of a given image. Curvelet was initiated in the continuous domain and then discretized for sampled data. Contourlet transform was stated directly in the discrete domain and then expanded in the continuous domain for more mathematical analysis. The main parts of the contourlet transform are the Laplacian pyramid (LP) and directional filter bank (DFB). Bandlet aims to provide an adaptive representation while edge detection is required, it is unreliable and noise sensitive as well.

Recently, several new transforms have been proposed such as the spectral graph wavelet transform [5] and the scattering transform [6]. The former designs a framework for constructing wavelets on arbitrary weighted graphs. By analogy with classical wavelet operators in the Fourier domain, it has been shown that scaling may be implemented in the spectral domain of the graph Laplacian. The spectral graph wavelets are localized in the small scale limit, and form a frame with easily calculable frame bounds. A scattering transform is a deep convolutional network with cascades of wavelet filters and average operators. The decomposition of one layer contains a low frequency part and several high frequency parts with different scales and directions. What is more, it calculates higher order coefficients by further iterating on high frequency parts of the last layer. It attains a translation invariant representation which is Lipschitz continuous to deformations. However, the inverse process is not available.

Motivated by these work, this paper constructs a new invertible deep convolutional network where the decomposition is realized in both the lower frequency and the higher frequency parts layer by layer. Thereby it provides a good performance on images with much detail information, e.g. textures. We directly explore it in the discrete-domain and give its construction with iterated directional filter banks [7]. Subject to this, perfect reconstruction is available if we set the kernel qincunx filters to be biorthogonal or orthogonal. A frequency recombinaltion progress is adjoined to the final layer of the convolution network. The manipulation transforms the uniform frequency partition to nonuniform one and thus increases the efficiency and flexibility of the network. Good results have also been achieved in numerical experiments of nonlinear approximation and denoising.

2. DISCRETRE-DOMAIN CONSTRUCTION

The underlying deep convolutional network architecture is shown in Fig. 1, which gives a brief overlook of whole operating procedure.

[*]The work has been partially supported by the NSFC grants 61425011, U1201255, 61271218, 61529101, and Shu Guang project (13SG13).

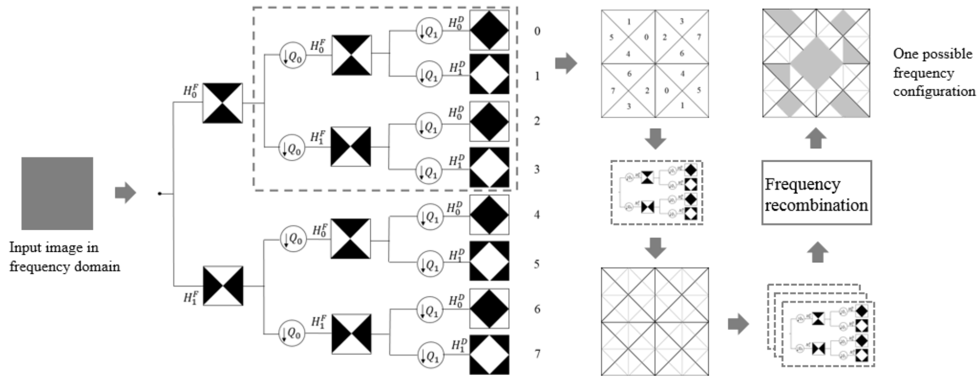


Fig. 1. The network architecture, which illustrates the frequency partition in each step. The operation in the dotted box is iterated before a frequency combination manipulation. The combinated areas are shown in gray color.

2.1. Multiresolution Directional Filter Bank

If a 2-D signal $x[n]$ where $n = [n_1, n_2]^T$ is downsampled by a decimation matrix M, then the discrete Fourier transform of the downsampled signal $x_d[u] = x[Mu]$ is

$$X_d(\omega) = \frac{1}{|M|} \sum_{k \in \mathcal{N}(M^{\mathrm{T}})} X(M^{-\mathrm{T}}\omega - 2\pi M^{-\mathrm{T}}k) \qquad (1)$$

where $\mathcal{N}(M)$ is defined as the set of integer vectors of the form Mt, where $t \in [0, 1)^2$. The number of elements in $\mathcal{N}(M)$ is equal to $|M|$. For an M-fold upsampling, the input $x(u)$ and the output $x_u(u)$ are related by

$$\begin{aligned} x_u[n] &= \begin{cases} x(M^{-1}n), & \text{if } n = Mk \\ 0, & \text{otherwise} \end{cases} \\ X_u(\omega) &= X(M^T\omega). \end{aligned} \qquad (2)$$

Bamberger and Simith introduced a 2-D directional filter bank (DFB) that could be maximally decimated while achieving perfect reconstruction [8]. Inspired by the DFB implementation via a l-level tree-structured decomposition, here we realize a uniform quincunx DFB whose subband configuration is shown in Fig. 1 by a binary tree structure. The decimated matrix and filters are given in Eq. (3) and Fig. 2. The frequency spectrum of the input signal can be splited into a lowpass and a highpass channel through a diamond filter pair, or into a horizontal and a vertical channel using a fan filter pair. In fact, the diamond filter can be fulfilled by simply modulating the fan filter. Combining the supports of these three filters, we get the desired support configuration.

$$Q_0 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad Q_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \qquad (3)$$

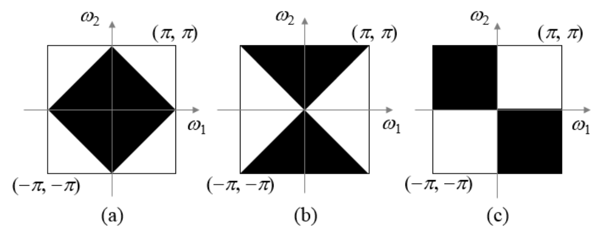


Fig. 2. Possible support configurations for the filters in the QFB. (a) Diamond filter. (b) Fan filter. (c) Quadrant filter

2.2. Architecture

2.2.1. Decomposition by Iterated Filter Banks

From Section 2.1, we know that an eight-directional frequency partition can obtained by a three-level decomposition. Notice the fact that the complete frequency square after three levels is uniformly divided into several triangular regions. It is a significant advantage since the four subblocks in four conners of the frequency square have the same frequency configuration type as the quincunx filter. Thus, further decompositon on both the lower frequency and high frequency parts can be achieved by directly repeating the process in the box of Fig. 1. Instead of iterating the same filter pair layer by layer which would lead to more directions but same scale, we iterating two kinds of filter pairs in turn so as to obtain finer divisions of the frequency square. Therefore fan filter pairs are adoped in even layers and diamond filter pairs are adopted in odd layers.

We index the channels of a DFB with l-levels from top to bottom as 0 to $2^l - 1$. The channel index k can be associated with a sequence of path types $(t_1, t_2, ..., t_l)$ of the filter banks from the second level leading to that channel by

$$k = \sum_{i=1}^{l} t_i 2^{l-i}$$

With this path type, using the equivalent filter banks, the sequence of filtering and downsampling for the channel k can be written as

$$\boxed{H_k^l(\omega)} \longrightarrow (\downarrow M_k^l) \tag{4}$$

where

$$H_k^l(\omega) = 2H_{t_1}^F \prod_{i=2}^{l-1} F_{t_i}, \quad \text{with } F_{t_i} = \begin{cases} H_{t_i}^F, & \text{if } i \text{ is even} \\ H_{t_i}^D, & \text{otherwise} \end{cases}$$

$$M_k^l = \prod_{i=1}^{l-1} Q_{(i+1)\bmod 2}$$

The synthesis system is symmetric to the analysis system. Similar to the analysis in Eq. (4), we can also transform the synthesis side into an upsampling operation by the overall sampling matrix M_k^l followed by an equivalent synthesis filter $G_k^l(\omega)$ to each channel. Then the reconstructed signal is

$$\hat{X}(\omega) = \sum_{k=1}^{2^l-1} G_k^l(\omega) \frac{1}{M_k^l} \sum_{k \in \mathcal{N}(M_k^{l^{-\mathrm{T}}})} H_i(\omega - 2\pi M_k^{l^{-\mathrm{T}}} k) \times X(\omega - 2\pi M_k^{l^{-\mathrm{T}}} k)$$

In fact, the reconstruction process is implemented level by level. It has been proven that the DFB provides a biorthogonal or orthogonal expansion if and only if its kernel quincunx filter banks are biorthogonal or orthogonal respectively [8]. It ensures a perfect reconstruction if the synthesis filters is the time-reversed version of the analysis filters.

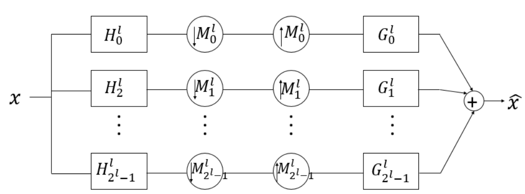


Fig. 3. The multi-channel view of a l-level DFB that has 2^l channels with equivalent filters and sampling matrics.

2.2.2. Frequency Recombination

The decomposition in convolutional network using iterated qincunx directional filter banks has been constructed in Section 2.2.1. The full channel decomposition scheme provides a full frequency partition. This scheme may not be useful and efficient when it comes to the practical scene since the large amount of subbands increases the computation cost without obvious promotion. To this point, we introduce a frequency recombination process following the decomposition process. Within various frequency recombination methods, the architecture layout in Fig. 1 gives an example. It attaches a strong flexibility to the whole architecture and leads to great potentials in various fields. Fig. 4 gives three examples of frequency recombination. The left one and the right two correspond to a three-level and a four-level decomposition networks. We can get finer and more abundant frequency configuration effect as the level increases. This is the reason why we construct a *deep* convolutional network. Imagining that if we set the network to be deep enough, an approximately continuous frequency domain partition can be achieved. In fact, the frequency recombination manipulation can be considered as another network since different paths exist and the recombination can also be in operation layer by layer.

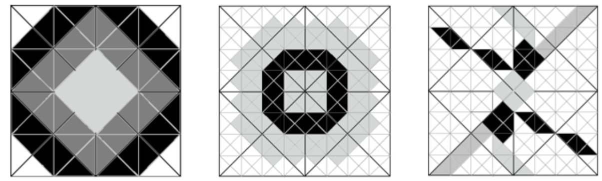


Fig. 4. Some examples of frequency recombination.

3. EXPERIMENTS

In this section, several experiments of nonlinear image approximation and denoising have been evaluated to help explore and demonstrate properties of the proposed scheme.

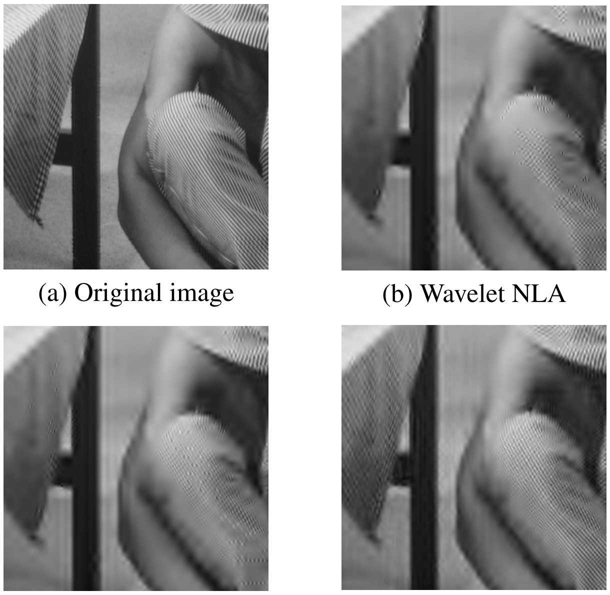
3.1. Nonlinear Approximation

First, we compare the nonlinear approximation (NLA) performance of the wavelet, contourlet transform and proposed network. In these NLA experiments, for a given value M, we select the M-most significant coefficients in each transform domain, and then compare the reconstructed images from these sets of M coefficients.

We choose six gray scale images in size of 512×512 or 256×256 for testing. We choose the 3% most significant coefficients for approximation on each image. The results are given in Table. 1. In five images, the proposed structure achieves the best performance among the three methods. Especially in *barbara* and *fingerprint*, there is a significant gain of 0.48 dB and 0.71 dB in peak signal-to-noise ratio (PSNR), respectively.

Fig. 5 shows a visual effect on the *barbara* image, where the PSNR values of the three reconstructed images are 26.03 dB, 26.46 dB and 26.92 dB, respectively. It indicates that the proposed network outperforms the other two transforms, especially the directional textures on cloth.

For more evidence, we repeat the nonlinear reconstruction process using different percentage of coefficients which ranges from 1.4 to 5.0. As shown in Fig. 6, the reconstructed images by the proposed network have consistently higher PSNR compared to wavelet and contourlet transforms. Another observation is that the gap between contourlet and the



(a) Original image (b) Wavelet NLA

(c) Contourlet NLA (d) Proposed NLA

Fig. 5. Nonlinear approximation by the wavelet, contourlet and proposed methods. (a) is the original image. (b), (c) and (d) show the reconstructed images from the 6553-most significant coefficients.

Table 1. PSNR of reconstructed image of different transforms

PSNR(dB)	wavelet	contourlet	proposed
barbara	26.03	26.46	**26.92**
cameraman	26.39	25.92	**26.58**
peppers	32.19	32.35	**32.56**
fingerprint	21.68	22.01	**22.72**
lena	**31.53**	31.01	30.60
baboon	22.03	22.09	**22.43**

proposed transform gets larger as the percentage of coefficients increases. It implies that the convolutional network has a more obvious superiority if more coefficients are involved.

3.2. Denoising

Similar to approximation, denoising can also be achieved based on keeping the most significant coefficients. A simple thresholding scheme on the decomposition coefficients is used in experiments to make denoising. Likewise, we use the same six test images and the results are given in Table. 2.

An observation is that the proposed method outperforms wavelet and contourlet transforms in general since the PSNR values rank top in five test images. Furthermore, more significant improvement is achieved in *barbara* and *fingerprint*. This strengthen the fact that our deep convolutional network has an especially great performance on images with much details.

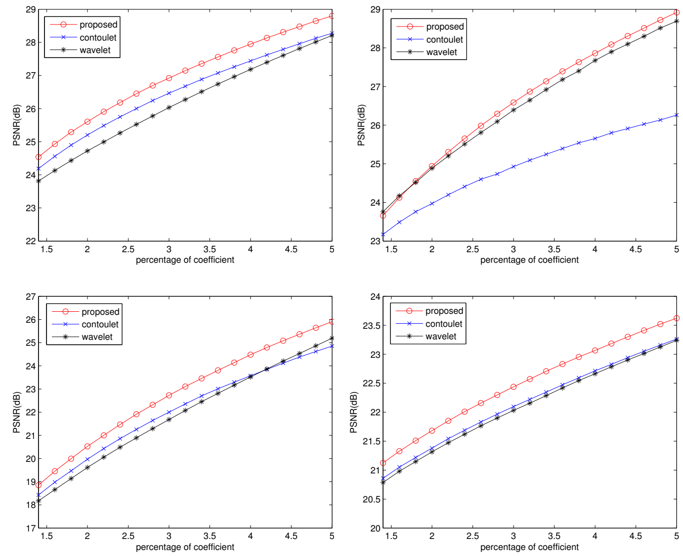


Fig. 6. Comparison of nonlinear approximation using wavelet, the contourlet and the proposed transform . A series of coefficients are chosen. Upper left: *barbara*. Upper right: *cameraman*. Bottom left: *fingerprint*. Bottom right: *baboon*

Table 2. Comparison of various methods in denoising

PSNR(dB)	noisy	wavelet	contourlet	proposed
barbara	22.95	24.84	25.12	**25.69**
cameraman	21.81	25.33	25.49	**28.41**
peppers	23.08	28.17	28.20	**32.56**
fingerprint	20.71	21.25	21.29	**22.76**
lena	24.04	28.22	**28.38**	28.23
baboon	21.65	22.54	22.81	**22.92**

4. CONCLUSIONS

In this work, we construct a new invertible deep convolutional network for image representation. What makes our network stand out of other methods mainly lies in two aspects. First, instead of iterating decomposition on every low frequency part and leaving the high frequency parts unchanged, we provide a more reasonable and flexible frequency partition scheme which enables frequency redistribution in various parts. The second contribution is its efficient and simple implementation by using iterated directional filter banks which allow an invertible access. Furthermore, by adjoining a frequency recombination process, the architecture indicates a promising future in various image processing applications.

5. REFERENCES

[1] S. Mallat, "A wavelet Tour of Signal Processing," 3rd ed. New York: Academic, 2008.

[2] E. Candes and D. Donoho "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219-266, Feb. 2004.

[3] M. Do, and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process*, vol. 14, no. 12, pp. 2091-2106, Dec. 2005.

[4] E. Pennec and S. Mallat "Sparse geometric image representation with bandelets," *IEEE Trans. Image Process*, vol. 14, no. 4, pp. 423-438, Apr. 2005.

[5] D. K. Hammond, P Vandergheynst, R Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis, Elsevier*, vol. 30, no. 2, pp. 129-150, 2011.

[6] J. Bruna and S. Mallat "Invariant scattering convolution networks," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 35, no. 8, pp. 1872-1886, 2013.

[7] M. Vetterli and C. Herley "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Process*, vol. 40, no. 9, pp. 22072232, Sep. 1992.

[8] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," in *IEEE Trans. Signal Process*, vol. 40, no. 4, pp. 882-893, Apr. 1992.